

# Why (model) predictions fail

2024-10-14

“It is difficult to make predictions, especially about the future.”

— Karl Kristian Steincke (1948) *Farvel Og Tak.* (Translated from Danish)

In my experience, people love to trash-talk weather forecasts<sup>1</sup>. *A 50% chance of rain? I don't see any clouds!* In actual fact, weather forecasts have gotten pretty darn good over the years! They can predict general weather patterns out for a week<sup>2</sup> and predict storms forming and their likely trajectory! **It's a miracle of science, computing, and international negotiations!** But no, rather than recognize the enormity of this success, we focus on the misses, the uncertainty, the wrong predictions.

Well I have a counter example: The stock market. You think meteorology is poor at prediction? Look at any economic prognosticator and you will see vague assertions of causal associations and very broad predictions (all wrapped in a veneer of confidence bordering on arrogance), yet most will still be wrong!<sup>3</sup> Or focus specifically on those trying to predict the stock market days or weeks into the future. They have essentially perfect, real-time data on all of the interactions (=trades), armies of smart people (“quants” = physics or math PhD’s) and huge incentives to get it right (=\$\$\$), but very, very few are any good (and just over very, very short periods) and most are beaten over the long run by a market averaged approach<sup>4</sup>. If they can’t get it right, why should we expect models of COVID-19 or fisheries or climate or, yes, even weather to perfectly predict years into the future?!

OK, OK, I'll step off my soap box and stop ranting. There is a good question here: why do models fail? I've quoted George Box in saying that all models are wrong, but some are useful. Let's focus on the question of why they are wrong.

I think we can break this into three types of issues.

## Parameterization: your inputs are problematic

As the adage goes, garbage in, garbage out. If you give a perfectly decent model bad inputs or parameters, you'll get bad or even silly results. In biology we often have to make reasonable guesses at parameter values or try to estimate them in sometimes very round-about way. Sometimes we're off...a lot. And that means that the model will give you unrealistic results. If you guess that  $R_0$  is 5 but the real value is closer to 1.5, then your model will predict a much, much bigger epidemic than you actually see.

The good news is that this is easy to deal with. First, we often create models, compare them to the data, and then refine them, tweaking parameters and so on. Models tend to produce better predictions as we give them better inputs. Second, if there is a parameter you know very little about, you can try a bunch of different values<sup>5</sup> over a range of possibilities and see how your model outputs change. If

<sup>1</sup> Also, everywhere I've lived I've heard someone say, “If you don't like the weather, just wait five minutes and it will change.” I'm in favor of a **moratorium on that phrase**.

<sup>2</sup> In most places. Here's an analysis of where it's predictable and where it is not.

<sup>3</sup> When asked about past misses, they usually spout some post-hoc explanation for how there was some unknown or unknowable factor at work, so they were right, even though they were wrong!

<sup>4</sup> You didn't, and shouldn't, ask me for investing advice, but here is mine: look to low cost index funds and ETFs.

<sup>5</sup> Thankfully computers do this all very fast with little fuss!

it changes very little, great! It probably does not matter a lot. If the results vary a lot as you change the value of the parameter, then you a) know it's an important parameter and b) have a good reason to go ask a funding agency to give you money to go try and pin it down.

Good modelers are good at this and will provide a sense of what parameters they have confidence in and which are uncertain, as well as what the consequences of that might be. Beware the modeling paper or report that does not address uncertainty!

### An aside on explaining prior observations vs. predicting the future

When we say “models predict” we often mean two somewhat different things. The first is that we often are trying to use models to make sense of or explain what has already happened. For instance, the cycles of measles outbreaks in pre-vaccine era United Kingdom have been the subject of numerous modeling efforts to try to understand what drove these cycles and related questions about the critical community size<sup>6</sup>, the role of schools in transmission, and the importance of movement between communities. The data are the same and the goal is to use a simple-but-not-too-simple model to describe or explain the observed data reasonably well. We might say, “the model predicted that communities smaller than 5000 should have only short-lived, sporadic epidemics,” but importantly, we do not mean *in the future*. The world has changed in so many ways since that era (e.g., widespread vaccination) that no one would expect these models to be any good at predicting measles dynamics in the future. They are a tool for research and understanding.

Sometimes, like weather models, we *are* trying to predict, or *forecast*, what will happen in the future. We may have a few months of data on, say, COVID-19 hospitalizations and we hope to use those and other data to construct a model that predicts what COVID-19 hospitalizations look like in the future. In a very real sense we construct a model to “predict” the past—essentially tweak the model and parameter values so that they explain the existing dynamics of COVID hospitalizations—in order to predict the future. The hope is that the past predicts the future<sup>7</sup>. If a model can “get it right” in the past, it would seem that it is capturing the essential elements to do a good job predicting forward.

<sup>6</sup> Essentially the population size needed for infections to persist

### Another aside: a trade-off inherent in prediction

There is an inherent statistical trade-off between explaining observed data and predicting the future. Think about it this way. Imagine you have an SEIR model of COVID-19 that gets the general dynamics of the epidemic in, say, Seattle. You feel pretty good about it, but you look closer and see that you are missing some of the lows and highs. Perhaps there is a surge in cases at one time and at another time cases do not rise as fast as the model expected. Can you tweak your model to better fit both of those features? Absolutely! Maybe you add another term or make transmission a parabolic function of weather so that it is highest when

<sup>7</sup> If your spidey senses are tingling, you are right! As I note below, this can be a fraught assumption in a rapidly evolving pandemic!

the weather is warm and dry and very low at other times. Does that make strong biological sense? No idea! But it works!

Now imagine you keep tweaking your model until it does an admirable job of describing the features of the prior data. However, when you forecast what should happen in the future you are likely to see your model predicts crazy, often completely unrealistic dynamics! What you have done is **over fit** your model, adding lots of complexity to your model to fit what amounted to stochastic, essentially random noise<sup>8</sup>. When you project from this overly complex model, those little added terms and complexities (e.g., transmission as a complex function of weather) can become rather important. For instance, that transmission term that varied with weather? It might make transmission unrealistically low (or high) during warmer summer months. You would have been better off with a simpler model that didn't fit the prior data *too* well.

This is called the “bias-variance” trade-off and it crops up whenever we use models for prediction. Fortunately, there are statistical approaches to find the sweet spot along the spectrum of complexity. The upshot for you is simply that a good model is not perfect, but that's a good thing!

## Structural: you're missing something important

In much the same way that the exponential model is missing the phenomenon of finite resources (as well as space, immigration and emigration, and interactions with other species), models of all sorts might be missing some important factor or process. An epidemic model that assumes individuals become immune and resistant to re-infection would certainly do a poor job of describing an epidemic for a parasite that does not elicit immune memory. A model that assumes homogeneous mixing would not do very well in predicting the spread of infection in sessile<sup>9</sup> corals or sea fans. The point is that the structure of a model should represent the important processes reasonably well. If it does not, it will fail to describe a system or predict the future dynamics.

The good news from the scientific point of view is that you have just learned something!<sup>10</sup> However you were thinking your system worked, you were wrong! Congratulations! Now you just need to sort out *how* you were wrong and what else is happening.

There is an interesting feedback when making predictive models. The models and results can *change the future* such that the model no longer applies. Many early models of COVID-19 dynamics, for instance, posited that there were no changes in contact rates; that is, they assumed no one did anything about the epidemic. They proved very useful in motivating many interventions<sup>11</sup>, but as the world changed, those original models were missing something important (e.g., drastically reduced movement and contact rates) and were therefore almost immediately wrong about the future. It's not their fault that the world changed out from under them; had no one done anything, they might well have gotten the dynamics of COVID-19 pretty spot on! Who knows?

<sup>8</sup> You can *under fit* a model, too, so it does not capture the essential features in the data. It, too, will be bad at prediction.

<sup>9</sup> Attached at its base and so not mobile.

<sup>10</sup> It is true that science progresses more rapidly from “failures” than “successes.”

<sup>11</sup> Remember “flatten the curve”?

## Scale and scope: your model does not do what you want it to do / your data suck

Models are constructed with a particular scope. For instance, an SEIR model might describe the dynamics of COVID-19 in Seattle or Pullman. Data on the incidence of new COVID-19 cases in either city could be used to both parameterize the model and as a reality check for it. So could data on probable COVID-19 hospitalizations, antibody prevalence, or similar. But if you were hoping to explain mortality rates overall, regardless of the cause, the model would be inappropriate at best and downright mislead at worst. It was not designed to describe the dynamics of all types of mortality, only that from COVID-19, so unless COVID-19 were the only or dominant source of mortality, your model would not do a good job of explaining or predicting total mortality rates.

There are other ways that there can be a mismatch between the scope or scale of the model and the data you are trying to predict or explain. One would be a mismatch in time. For instance, our hypothetical COVID-19 model would be useless in trying to explain changes in the size of the US population from the census. First, as before, the model is meant to describe *epidemic* processes, not demographic changes in immigration and emigration, birth and death rates, and so on. Second, the census is conducted every decade whereas the COVID-19 pandemic (before it became endemic) was a much shorter event that might leave a small signature impact on population size, but would not explain longer-term dynamics<sup>12</sup>. There might also be spatial mismatches. Our model might be applied neatly to a city, but if data are reported only at a state or national level, well, they will not fit very well<sup>13</sup>.

The answer to this problem is, well, context dependent. Maybe you can find better data with the appropriate measure at the right spatial and temporal scale. Or maybe you can think of a way to increase or decrease the scope of your model. Or maybe you just have a model that shows what could happen or how things might work and then you try to draw out all of the implications of the processes in your model and see if they are logically consistent with the state of the world<sup>14</sup>. In any case, beware of extrapolating beyond the context or scope of the model<sup>15</sup>!

## Final thoughts

- When we *are* interested in predictions, we should think about the scope of the model. Sure, a weather model might not be great at telling you whether you will need an umbrella on a given day, but it's probably pretty good for the area as a whole. A model of average Malaria risk for a region in Africa might not do a good job of predicting risk at a particular village, but then again, it wasn't meant to!
- Similarly, we should be better about allowing for uncertainty in forecasts. If a forecast says a 40% chance of rain, that means there is a 40% probability of at least a quarter of a millimeter in that particular place during the forecast. That

<sup>12</sup> You might think this is all a bit ridiculous, but I assure you similar mismatches are not uncommon in the various sciences. You might be trying to study the effect of some small-ish process, but you only have very large or coarse data and so you try to smoosh them together and ignore all the ways that feels wrong... Not to pick on economics, but it seems rife with these awful, discordant models and data! (OK, I lied. I like to pick on economics!)

<sup>13</sup> Ask me about models of Lyme disease sometime if you want an ear-full!

<sup>14</sup> Or you find a suitably low-tier journal that will let this one slip past, or you wave your hands rapidly while you subtly change the nature of the argument, or ...



means that in 40 of 100 forecasts with a 40% chance of rain there will be rain ( $> 0.25\text{mm}$ ), but in the other 60 there will not be. The fact that it does not rain does not mean the forecast is right (or wrong). That was only one “roll of the dice,” so to speak. An elevated risk of COVID-19 infections does not necessarily mean that *you* will be infected (nor that you will not be infected); it applies to the population as a whole.

- It is worth remembering that models can be terribly useful, even if they are not in the business of prediction! Discrete-time (and thus potentially chaotic) logistic growth, Lotka-Volterra predator-prey and competition, metapopulation dynamics, island biogeography, and optimal foraging models are all examples of hugely influential models in ecology that are not intended to predict what happens in any particular system. Instead, they illustrate how particular forces or processes lead to particular outcomes. They help us think through the consequences of different scenarios. Similarly, a simple SIR model may not describe a real-world pandemic very well, but it can be very useful in thinking about which interventions might be the most effective.
- This leads to my final take-home message: models are often most useful for making *projections* rather than predictions; running scenarios meant to illustrate what would happen *if* such and such were true more than the precise level of some response (e.g., COVID-19 cases) at some point in time. That is, they are perhaps most useful for seeing the consequences of different actions (or inaction), assumptions, or causal structures. Sure, they are simplifications of reality and thus, by definition, wrong, but they can be wonderfully useful tools, even when “wrong.”